Could less be more when assessing patient-rated outcome in spinal stenosis?

Mannion, Anne F ; Fekete, Tamas F ; Wertli, Maria M ; Mattle, Michele ; Nauer, Selina ; Kleinstück, Frank S ; Jeszenszky, Dezsö ; Haschtmann, Daniel ; Becker, Hans-Jürgen ; Porchet, François

Abstract: STUDY DESIGN: Longitudinal study of the measurement properties of a brief outcome instrument. OBJECTIVE: In patients undergoing surgery for lumbar spinal stenosis, we compared the responsiveness of the Core Outcome Measures Index (COMI) with that of the condition-specific Swiss Spinal Stenosis Measure (SSM), an instrument developed to assess patients with neurogenic claudication. SUMMARY OF BACKGROUND DATA: The COMI is a validated multidimensional questionnaire for assessing the key outcomes of importance to patients with back problems. Being brief, it is associated with minimal respondent burden and high completion rates. However, for a given pathology, intuitively it may be expected to be less responsive than a condition-specific instrument. METHODS: A total of 91 patients (73±8 yr; 53% males) completed the following questionnaires before surgery: COMI, SSM, Roland Morris Disability Questionnaire, back trouble "Feeling Thermometer," pain numeric rating scale, EuroQoL-visual analogue scale. Twelve months postoperatively, 78/91 (86%) completed all the questionnaires again; they also rated the "global treatment outcome" (GTO; rated 1-5) and SSM "satisfaction with treatment result" (SSM-sat; rated 1-4), which were used as external criteria of treatment success. RESULTS: Scores for the external criteria of success (GTO/SSM-sat) correlated with the change scores (baseline to 12 mo) in COMI (r=0.57) and SSM (r=0.54) to a similar extent. Using receiver operating characteristics, with GTO or SSM-sat dichotomized as external criterion, the area under the curve was similar for the COMI change score (0.86-0.90) and the SSM (sub)scales (0.80-0.90). CONCLUSION: With either SSM-sat or GTO serving as the external criterion, COMI was as responsive as the SSM. The COMI is well able to detect important change in lumbar spinal stenosis and has the added benefit of reducing the response burden for the patient and facilitating outcome comparisons with other spinal pathologies. LEVEL OF EVIDENCE: 2.

DOI: https://doi.org/10.1097/BRS.0000000000000751

Spine

SPINE Volume 40, Number 10, pp 710-718

CLINICAL CASE SERIES

Could Less Be More When Assessing Patient-Rated Outcome in Spinal Stenosis?

Anne F. Mannion, PhD,* Tamas F. Fekete, MD,* Maria M. Wertli, MD,† Michele Mattle, MSc,† Selina Nauer, MSc,* Frank S. Kleinstück, MD,* Dezsö Jeszenszky, MD, PhD,* Daniel Haschtmann, MD,* Hans-Jürgen Becker, MD,* François Porchet, MD,* and Lumbar Spinal Stenosis Outcome Study (LSOS) Group

Study Design. Longitudinal study of the measurement properties of a brief outcome instrument.

Objective. In patients undergoing surgery for lumbar spinal stenosis, we compared the responsiveness of the Core Outcome Measures Index (COMI) with that of the condition-specific Swiss Spinal Stenosis Measure (SSM), an instrument developed to assess patients with neurogenic claudication.

Summary of Background Data. The COMI is a validated multidimensional questionnaire for assessing the key outcomes of importance to patients with back problems. Being brief, it is associated with minimal respondent burden and high completion rates. However, for a given pathology, intuitively it may be expected to be less responsive than a condition-specific instrument.

Methods. A total of 91 patients (73 ± 8 yr; 53% males) completed the following questionnaires before surgery: COMI, SSM, Roland Morris Disability Questionnaire, back trouble "Feeling Thermometer," pain numeric rating scale, EuroQoL-visual analogue scale. Twelve months postoperatively, 78/91 (86%) completed all the questionnaires again; they also rated the "global treatment outcome" (GTO; rated 1–5) and SSM "satisfaction with treatment result" (SSM-sat; rated 1–4), which were used as external criteria of treatment success.

Results. Scores for the external criteria of success (GTO/SSM-sat) correlated with the change scores (baseline to 12 mo) in COMI ($r = 0.57$) and SSM ($r = 0.54$) to a similar extent. Using receiver operating characteristics, with GTO or SSM-sat dichotomized as external criterion, the area under the curve was similar for the COMI change score (0.86–0.90) and the SSM (sub)scales (0.80–0.90).

Conclusion. With either SSM-sat or GTO serving as the external criterion, COMI was as responsive as the SSM. The COMI is well able to detect important change in lumbar spinal stenosis and has the added benefit of reducing the response burden for the patient and facilitating outcome comparisons with other spinal pathologies.

Key words: spinal stenosis, outcome, Swiss Spinal Stenosis Measure (SSM), Zurich Claudication Questionnaire, Core Outcome Measures Index (COMI), responsiveness, validity, receiver operating characteristics, registries.

Level of Evidence: 2

Spine 2015;40:710–718

In judging the outcome of spine surgery, it is now accepted that the focus should be firmly placed on the patient's perspective, assessing factors of importance to them such as symptoms, function, and quality of life. However, the availability of many different instruments for each of these domains, and the lack of their standardized use, compromises meaningful comparison among different diagnostic groups, treatment procedures, and study groups. In recognition of this problem, in 1998, a seminal article was published by Deyo *et al*,[1] in which recommendations were made for a standardized set of outcome measures for use in patients with back pain. There was general consensus that the most appropriate core outcome measures should include the domains pain, back-specific function, generic health status (well-being), work disability, social disability, and patient satisfaction.[1,2] Accordingly, the group proposed a parsimonious set of 6 questions that would cover each of these domains yet be brief enough to be practical for routine clinical use, quality management, and possibly also more formal research studies.[1] The psychometric characteristics of the core set were subsequently examined in both surgical and conservative patients with back pain, and the reliability, validity, and sensitivity to change of the individual core questions and of a "multidimensional sum score" was established.[3,4] Another single question was added to the core set to assess "overall quality of life" (taken from the World Health Organisation Quality of Life BREF [WHO-QoL BREF] questionnaire) because this domain seemed to be delivering different information to the (symptom-specific) "overall well-being" question in the original core set.[3] This group of questions formed the basis of what is now known as the Core Outcome Measures Index (COMI),[5] subsequently validated as an instrument in itself

From the *Spine Center, Schulthess Klinik, Zurich, Switzerland; and †Horten Centre, University Hospital Zurich, Zurich, Switzerland.

Acknowledgment date: July 31, 2014. Revision date: October 30, 2014. Acceptance date: December 10, 2014.

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work.

Relevant financial activities outside the submitted work: consultancy, royalties.

Address correspondence and reprint requests to Anne F. Mannion, PhD, Spine Center, Schulthess Klinik, Lengghalde 2, 8008 Zurich, Switzerland; E-mail: anne.mannion@yahoo.com

DOI: 10.1097/BRS.0000000000000751

and adapted for use in many different languages.[3,4,6–11] The COMI has since been incorporated into the Eurospine Spine Tango registry as their outcome measure of choice[12–14] and has been shown to be feasible to implement on a prospective basis for all patients undergoing surgery and long-term follow-up[15] within a busy (>1000 cases/yr) tertiary care Spine Unit.[5]

With just 1 question per domain, the COMI enables the efficient assessment of large numbers of patients, with minimal respondent burden. However, intuitively, it may be expected to be less responsive than a condition-specific instrument when evaluating the outcome of treatment for a specific pathology. Responsiveness, that is the ability to measure meaningful change in a clinical state, is one of the most important properties of an outcome instrument used for evaluative purposes (*i.e.*, longitudinal assessment).[16,17] Generally speaking, condition-specific instruments are expected to be highly responsive because all their items are specifically targeted toward addressing the condition in question. Previous studies have shown, for example, that the Swiss Spinal Stenosis Measure (SSM) is more responsive than either the Sickness Impact Profile or its derivative the Roland Morris Disability Questionnaire (RMDQ) in assessing patients with spinal stenosis.[18]

The aim of this study was to compare the performance of the COMI (in terms of its internal and external responsiveness) with that of the widely accepted condition-specific SSM and other spine outcome instruments, in patients undergoing surgery for lumbar spinal stenosis.

MATERIALS AND METHODS

Patients

Patients (N = 91; Table 1) were from a single center, in the surgical arm of a multicenter prospective cohort study evaluating outcomes after surgical and conservative treatment of central spinal stenosis (explained in detail in Steurer *et al*[19]). The main inclusion criteria were age more than 50 years; uni-/bilateral neurogenic claudication; verified diagnosis of spinal stenosis upon imaging; and ability to complete questionnaires in German. The main exclusion criteria were cauda equina syndrome; current fracture, infection, or significant deformity (>15° lumbar scoliosis); and clinically relevant peripheral arterial disease.

All patients were simultaneously registered in the local center's in-house Outcomes Database, nested within the Eurospine Spine Tango data acquisition system. Patients underwent either simple decompression (71%) or decompression with instrumented fusion (29%).

Questionnaires

The validated German versions of the questionnaires described in Table 2 were completed by the patients preoperatively and 12 months postoperatively. The COMI is shown in Figure 1.

At 12 months postoperatively, an additional question enquiring about the global treatment outcome (GTO)[5] was administered with the COMI: "overall, how much did the operation help your back problem?", with 5 response categories (1, helped a lot; 2, helped; 3, helped only little; 4, did not help; and 5, made things

TABLE 1. Baseline Patient Demographic and Clinical Data (N = 91)

	Mean (SD) or %
Age (yr)	72.9 (7.5)
Weight (kg)	78 (14)
Sex	
Male, N (%)	48 (53)
Female, N (%)	43 (47)
Height (cm)	169 (9)
BMI (kg × m^{-2})	27.3 (4.2)
Low back pain (Yes)	81%
Leg/buttock pain intensity (0–10)	4.4 (2.8)
Back pain intensity	6.4 (2.4)
Leg pain minus back pain intensity	2.1 (3.3)
Duration of problem (%)	
<3 mo	4
3–6 mo	12
6–12 mo	11
>12 mo	73
Comorbidity	
Cumulative illness rating scale (0–56)	9.7 (3.3)
ASA (%)	
I	12
II	57
III	31
Living conditions (%)	
Alone	29
With others (partner, family)	71
Highest education (%)	
Compulsory schooling	21
Apprenticeship/technical college	63
University	16
Occupation (%)	
Full-time work	3
Part-time work	9
Housewife	4
Retired	84
Marital status (%)	
Married	69
Divorced/separated	11
Widowed	13
Single	7

SD indicates standard deviation; ASA, American Society of Anesthesiologists.

TABLE 2. Questionnaires Administered Preoperatively and at 12 Months Postoperatively

Instrument	Brief Description	Time-Frame	Scoring
COMI[5]	Includes items for the domains: pain intensity (back and leg/buttock pain), back-related function, symptom-specific well-being, general quality of life, and social disability and work disability (Figure 1).	All items refer to "the last wk" except disability (last 4 wks)	COMI**:** 0 (best)–10 (worst)
SSM[18] (also known as the Zurich Claudication Questionnaire and the Brigham Spinal Stenosis Questionnaire)	3 main subscales: SSM-symptoms subscale, comprising 7 items on pain (severity, frequency, back pain, leg pain, numbness, weakness, and balance) disturbance This scale can be further split into the SSM-Pain subscale (containing severity, frequency, back pain) and SSM-Neuroischemic subscale (containing leg pain, numbness, weakness, balance disturbance) SSM-Physical Function subscale, comprising 5 items on: walking distance and pain on walking outdoors for pleasure, while shopping, around the house, and from bedroom to bathroom SSM-Satisfaction subscale comprising 6 items on satisfaction with various outcomes after treatment (used postoperatively only). SSM-average gives an unweighted average score for the symptom and function scales	Past mo	SSM-symptoms: 1 (best)–5 (worst) SSM-pain and SSM-neuroischemia: 1 (best)–5 (worst) SSM-physical function: 1 (best)–4 (worst) SSM-satisfaction: 1 (best)–4 (worst) SSM-average: 1 (best)–4.5 (worst)
RMDQ[38,39]	Enquires as to whether the back problem hinders the performance of 24 activities of daily living, with possible responses of "Yes" and "No"	Today	0 (best)–24 (worst)
EQ-VAS[40]	Records the respondent's self-rated health on a vertical VAS where the endpoints are labeled "best imaginable health state" (100) and "worst imaginable health state" (0)	Today	0 (worst)–100 (best)
Back-trouble Feeling Thermometer	A vertical VAS measuring the severity of back-trouble/problems, where the endpoints are labeled "extremely severe" (100) and "none" (0)	Past wk	0 (best)–100 (worst)
Pain 11-point NRS[41]	Pain intensity rated by selecting a number from 0–10 from a line of 11 boxes in which each of these numbers is written. The endpoints are "no pain" (0) and "worst pain imaginable" (10)	Past wk	0 (best)–10 (worst)

The COMI was presented separately (as routine follow-up) from the other questionnaires, which were administered as part of the cohort study.[19] As long as the 2 questionnaire booklets were completed within 1 mo of another, their data were included in the analysis.
COMI indicates Core Outcome Measures Index; SSM, Swiss Spinal Stenosis Measure; RMDQ, Roland Morris Disability Questionnaire; VAS, visual analogue scale; NRS, numeric rating scale.

worse). This was used as an external criterion of success, with the scores being dichotomized into "successful" (1 and 2) and "not successful" (3, 4, and 5)[20] for some of the subsequent analyses. Item 1 on the 12-month SSM-satisfaction scale (SSM-sat) served as a second external criterion of success: "How satisfied were you with the overall result of the operation?" with 4 response categories (1, very satisfied; 2, satisfied; 3, dissatisfied; and 4, very dissatisfied). These scores were also dichotomized to form "successful" (1 and 2) and "not successful" (3 and 4) groups for some subsequent analyses.

Statistical Analyses

On the basis of previous recommendations,[21] we aimed for a sample size of 50 to 100 patients (by 12 mo follow-up). Parametric statistics were employed: though individual items were measured on ordinal scales, the data for the instrument scores were not heavily skewed and could hence be treated as interval without introducing notable bias.[22] Data are presented as means and standard deviations (SDs) unless otherwise stated.[21–23]

The "internal responsiveness"[24] of the instruments was given by their corresponding standardized response means (mean change score from baseline to 12 mo/SD of the change score).

The "external responsiveness"[21,24] (or "longitudinal validity"[21]) of the COMI was evaluated by the strength of the correlation between its change scores (preoperatively to 12 mo postoperatively) and those of the SSM subscales and other outcome instruments. The correlation between these change

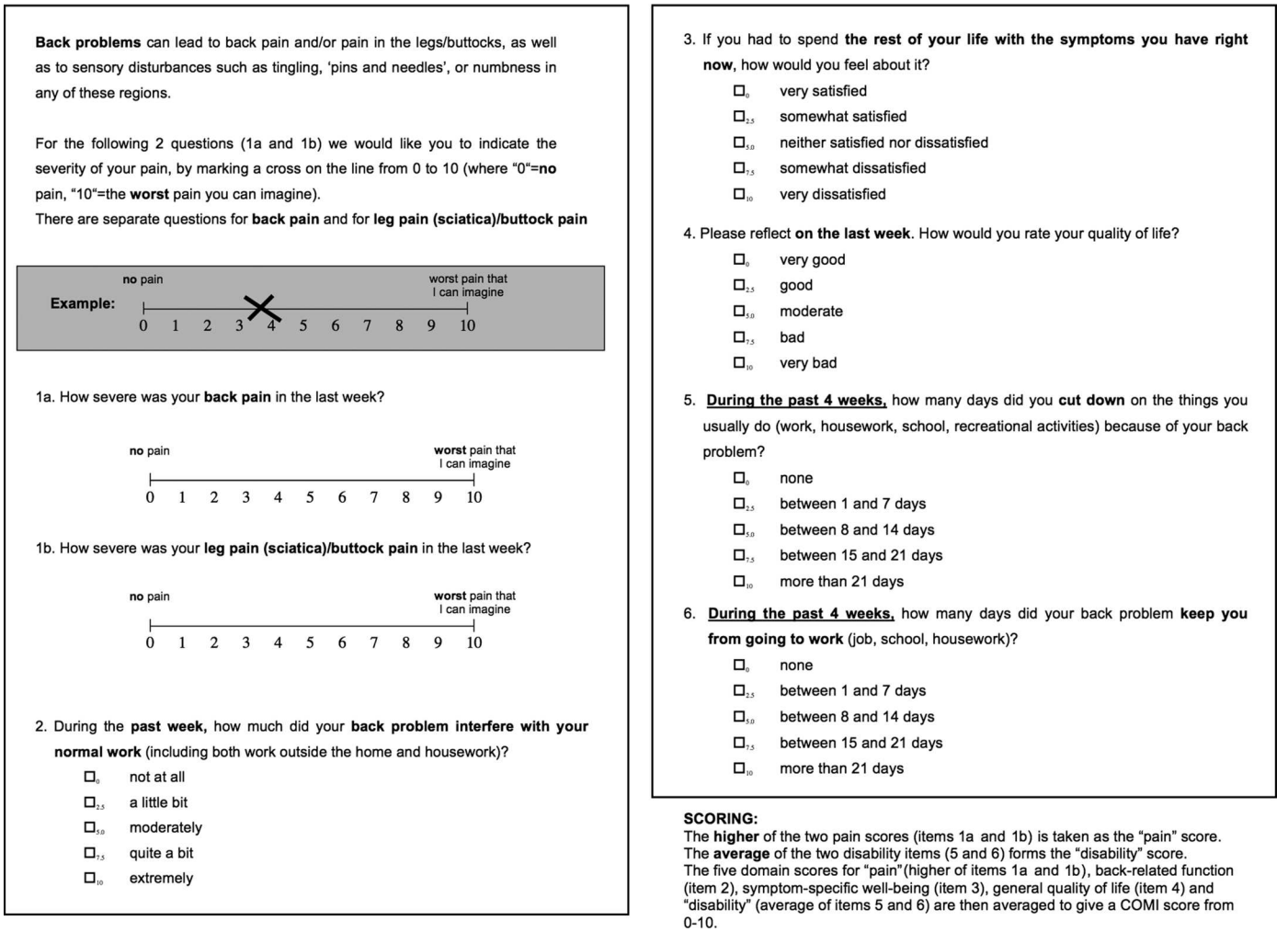

Figure 1. Core Outcome Measures Index (COMI).

scores and the GTO and the SSM-sat scores 12 months postoperatively (using the scores on the 5- and 4-point scales, respectively) was also examined. It was hypothesized that correlation coefficients for all these relationships, if the variables were measuring similar attributes, would range from 0.4 to 0.8.[22] External responsiveness was also assessed using receiver operating characteristics (ROC) curves. In this context, responsiveness was described in terms of the probability of the change scores correctly classifying patients who were successful (sensitivity) and not successful (specificity) on the external criterion.[24] The area under the ROC curve (AUC; with exact binomial confidence intervals) indicated the probability of correctly discriminating between "successful" and "unsuccessful" patients based on the change in instrument score. An AUC of 0.5 indicates discrimination no better than chance and an AUC of 1.0 indicates perfect discrimination (100% sensitivity and 100% specificity); an AUC > 0.7 reflects adequate responsiveness.[21] Logistic regression analyses were carried out, in addition, to evaluate the AUC for a combination of SSM subscale scores as independent variables with GTO and SSM-sat as dependent variables.

The analyses were carried out using Statview 5.0 (SAS Institute Inc., San Francisco, CA) and Medcalc (MedCalc Statistical Software, Mariakerke, Belgium). $P < 0.05$ were considered to be statistically significant.

RESULTS

Twelve months postoperatively, 90/91 (99%) patients returned the COMI and 79/91 (87%) returned the questionnaire booklet containing the SSM. Missing data were rare (see footnotes to Table 3).

Responsiveness

Table 3 shows the mean scores for each instrument preoperatively and at 12 months postoperatively, as well as the standardized response mean for these change scores. There was a significant ($P < 0.0001$) improvement in all instrument scores from baseline to 12 months postoperatively. The SRM was highest for SSM-average (and SSM-pain) (1.50), followed by COMI (1.44), SSM-symptoms (1.43), the Feeling Thermometer (1.39), the numeric rating scale (1.28), RMDQ (1.13), and SSM-physical function (1.00). The EuroQol-visual analogue scale showed the lowest SRM (0.55) (Table 3).

Table 4 shows the inter-relationships between the instruments' change scores (preoperatively to 12 mo postoperatively) and between each of these and the ratings on the 4- or

TABLE 3. Preoperative, 12 mo Follow-up, and Change Scores for the COMI and the Comparator Instruments

	Baseline Score		12 mo postoperatively			Change Score From Preoperatively to 12 mo Postoperatively		
	Mean	SD	Mean	SD	ANOVA *P*	Mean	SD	SRM pr–12 mo
COMI	6.8	1.8	2.9	2.4	<0.0001	3.9	2.7	1.44
SSM-symptoms	3.1	0.6	2.0	0.8	<0.0001	1.0	0.7	1.43
*SSM-pain	3.6	0.6	2.2	0.9	<0.0001	1.5	1.0	1.50
*SSM-neuro	2.7	0.8	1.9	0.8	<0.0001	0.7	0.8	0.89
SSM-physical function	2.2	0.6	1.5	0.5	<0.0001	0.7	0.7	1.00
SSM-average	2.6	0.5	1.7	0.6	<0.0001	0.9	0.6	1.50
RMDQ	12.3	5.2	7.0	5.4	<0.0001	5.4	4.8	1.13
EQ-VAS	56.7	21.9	73.4	22.1	<0.0001	16.7	30.5	0.55
Feeling Thermometer	65.5	19.7	28.1	24.3	<0.0001	37.4	27.0	1.39
NRS	6.4	1.9	2.7	2.5	<0.0001	3.7	2.9	1.28

Data are from N = 77 to N = 79. In addition to missing baseline data (1 COMI (administrative error); 2 patients' EQ-VAS; 2 patients' Feeling Thermometer; in 1 patient, 3 SSM-symptom items missing) 1 patient returned no questionnaires at all at follow-up and 11 patients failed to complete the questionnaire booklet containing the SSM, RMDQ, EQ-VAS, Feeling Thermometer, and NRS (either they did not respond to contact or actively withdrew because of the burden of participation). Furthermore, in 1 patient, just the EQ-VAS was not completed.

**Further subdomains of the SSM-symptoms subdomain (see Table 2 for details).*

COMI indicates Core Outcome Measures Index; SSM, Swiss Spinal Stenosis Measure; SRM, standardized response mean; SSM-neuro, SSM-Neuroischemic subscale; RMDQ, Roland Morris Disability Questionnaire; EQ-VAS, EuroQol visual analogue scale; NRS, numeric rating scale; ANOVA, analysis of variance.

TABLE 4. Correlation Matrix Showing the Inter-Relationships Between the Instruments' Change Scores (Preoperatively to 12 mo Postoperatively), Global Outcome, and Satisfaction (at 12 mo Postoperatively)

Change Scores (or for GTO and SSM-Sat, Score at 12 mo Postoperatively)	COMI	SSM-Symptoms	*SSM-Pain	*SSM-Neuro	SSM-Physical Function	SSM-Average	Roland Morris	EQ-VAS	Feeling Thermometer	NRS	GTO
COMI	1.00										
SSM-symptoms	0.58	1.00									
*SSM-pain	0.70	0.81	1.00								
*SSM-neuro	0.29	0.86	0.39	1.00							
SSM-physical function	0.55	0.67	0.67	0.46	1.00						
SSM-average	0.62	0.92	0.81	0.73	0.91	1.00					
RMDQ	0.43	0.35	0.32	0.28	0.39	0.41	1.00				
EQ-VAS	−0.33	−0.25	−0.26	**−0.16**	**−0.21**	−0.25	**−0.18**	1.00			
Feeling Thermometer	0.63	0.58	0.57	0.40	0.51	0.60	0.43	−0.29	1.00		
NRS	0.64	0.64	0.63	0.45	0.55	0.65	0.43	−0.24	0.89	1.00	
GTO	−0.57	−0.49	−0.48	−0.35	−0.49	−0.54	−0.32	**0.11**	−0.48	−0.49	1.00
SSM-satisfaction	−0.57	−0.52	−0.51	−0.37	−0.47	−0.54	−0.46	**0.22**	−0.59	−0.55	0.79

All coefficients $P < 0.05$, except those in boldface (for which $r \leq 0.22$).

**Further subdomains of the SSM-symptoms subdomain (see Table 2 for details).*

COMI indicates Core Outcome Measures Index; SSM, Swiss Spinal Stenosis Measure; SSM-neuro, SSM-Neuroischemic subscale; RMDQ, Roland Morris Disability Questionnaire; EQ-VAS, EuroQol visual analogue scale; NRS, numeric rating scale; GTO, global treatment outcome.

TABLE 5. Results of ROC Analysis Using GTO and SSM-Satisfaction as the External Criteria for a "Successful" Outcome

	GTO		SSM-Satisfaction	
Instrument	**AUC**	**95% Exact Binomial CI**	**AUC**	**95% Exact Binomial CI**
COMI	0.863*	0.766–0.931	0.902*†	0.813–0.958
SSM-symptoms	0.816	0.711–0.895	0.873	0.778–0.938
SSM-pain	0.808	0.703–0.889	0.872	0.777–0.937
SSM-neuro	0.732	0.619–0.827	0.786	0.678–0.871
SSM-physical function	0.797	0.690–0.880	0.861	0.763–0.929
SSM-average	0.832	0.730–0.908	0.898	0.807–0.955
RMDQ	**0.631**	**0.514–0.739**	**0.688**	**0.572–0.788**
EQ-VAS	**0.644**	**0.527–0.750**	**0.634**	**0.516–0.741**
Feeling Thermometer	0.794	0.687–0.878	0.948*†	0.873–0.986
NRS	0.798	0.691–0.881	0.908*†	0.820–0.962
‡SSM-symptoms and function	0.840	0.739–0.914	0.893	0.802–0.952
‡SSM-pain, neuro, and function	0.825	0.721–0.902	0.894	0.803–0.953

All AUCs, $P < 0.05$ except for: RMDQ and EQ-VAS with either GTO or SSM-sat as external criterion (shown in boldface).

**$P < 0.05$ higher AUC than for RMDQ.*

†$P < 0.05$ higher AUC than for EQ-VAS.

‡AUC determined from logistic regression analysis, with these variables as independent variables and GTO or SSM-sat as dependent variable.

ROC indicates receiver operating characteristic; COMI, Core Outcome Measures Index; SSM, Swiss Spinal Stenosis Measure; SSM-neuro, SSM-Neuroischemic subscale; RMDQ, Roland Morris Disability Questionnaire; EQ-VAS, EuroQol visual analogue scale; NRS, numeric rating scale; AUC, area under the ROC curve; GTO, global treatment outcome; CI, confidence interval.

5-point scales for the external criteria SSM-sat and GTO. The change scores for COMI showed moderate, statistically significant correlations with those of the SSM-average and the SSM main subscales (r = 0.5–0.6; Table 4), within the hypothesized range (r = 0.4–0.8). The external criterion GTO correlated to a similar extent with the change score for COMI ($r = -0.57$) as with the change scores for SSM-symptoms ($r = -0.49$), SSM-function ($r = -0.49$), and SSM-average ($r = -0.54$). Similar results were found when SSM-sat was the external criterion.

Sixty-eight of 78 (87%) patients had a "good" outcome (were successful) according to GTO, and 70/78 (90%) according to SSM-sat. The results of the ROC analyses are shown in Table 5 and Figure 2. The AUCs were similarly high (>0.83) for COMI and SSM-average (or for the combined subscales in logistic regression), regardless of whether GTO or SSM-sat was used as the external criterion. This showed that the COMI had good discriminative ability. The AUCs for RMDQ and EuroQol-visual analogue scale were significantly ($P < 0.05$) lower than those of the COMI.

DISCUSSION

There is increasing focus on the use of patient-oriented measures for evaluating treatment outcome. Brief questionnaires are ideal for assessments that are to be made repeatedly over time[25] and they offer numerous advantages over longer instruments, including easier administration, higher completion rates,[26,27] lower respondent burden, lower costs of data collection/scoring, and easier score interpretation.[28] However, it is essential that they also display adequate psychometric (or

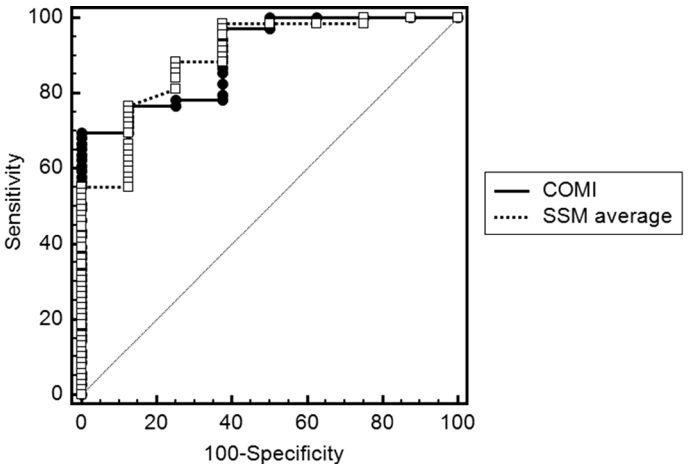

Figure 2. ROC curves for change scores for COMI (solid lines) and SSM average (dotted lines), using SSM-satisfaction as the external criteria for a "successful" outcome. ROC indicates receiver operating characteristic; COMI, Core Outcome Measures Index; SSM, Swiss Spinal Stenosis Measure.

"measurement") properties,[29] including the ability to detect important change resulting from treatment.[16] This study compared the performance of a brief, multidimensional outcome instrument (COMI) with that of the widely accepted condition-specific SSM in patients undergoing surgery for lumbar spinal stenosis. The COMI demonstrated a level of responsiveness comparable with that of the SSM.

Responsiveness was assessed in different ways.[24] Internal responsiveness was given by the standardized response mean (SRM, "group effect size").[24] This is a commonly used method,[24] but it fails to reveal whether the instrument also shows change where none is actually perceived by the patient: a group may improve on average, with a large effect size, but that average may comprise individual improvement, stability, or deterioration.[28] External responsiveness was given by the correlation between change scores (preoperatively to 12 months postoperatively) for the COMI and change scores for the SSM (sub)scales, as well as the correlation between each of these and the external criteria of treatment success; it was further assessed using ROC analyses. These analyses revealed the COMI to be one of the most discriminating outcome measures. The accuracy (or discriminatory ability) of the COMI change score to predict treatment success was 0.86 to 0.90, compared with 0.83 to 0.90 for the SSM-average or the SSM subscales combined (using logistic regression). The RMDQ had only moderate responsiveness (AUC, 0.63–0.69; $P > 0.05$), perhaps because its predominant focus is on activities typically limited by back pain. The low responsiveness of the EuroQol-visual analogue scale was not particularly unexpected, given that it is a generic rather than back-specific measure.[25]

It is generally considered that the more specific the items are to a condition, the more responsive will be the instrument. The SSM, also known as the "Zurich Claudication Questionnaire," contains 5 questions that assess walking ability and pain on walking, and it was hence expected to be highly specific for the patients in this study with neurogenic claudication. Ironically, the nonspecific nature of the questions in the COMI may have conferred upon it some of the positive characteristics of patient-specific/individualized questionnaires, rendering it highly responsive.[30] For example, the COMI's items on back function and disability simply enquire about interference with "normal work"/"things one normally does" rather than listing specific activities (that the patient may or may not normally do), allowing these items to be interpreted in relation to the patient's own particular difficulties. Similarly, the symptom-specific well-being item refers to the "symptoms the patient has now," without specifying any particular kind, and with the patient likely considering the ones most troublesome to them. This ensures that the items are always relevant to the given patient, making them more likely to be "shifting" items (*i.e.*, susceptible to change)[25] with effective treatment. Importantly, the COMI also uses the higher of the two pain scores (leg or back pain), rather than an average of the two, in forming the summary score. Patients with stenosis typically have both back pain and leg pain, but there is interindividual variation in the intensity of each. In this study, the leg pain minus low back pain intensity was 2 points on average, but with a SD of 3 points (Table 1). Clearly, then, there are patients with severe leg pain and no back pain, and some with only little/moderate leg pain and intense back pain. Pain is known to be one of the most responsive items in spinal surgery,[31] and if the effect of intense pain in the most painful region is "diluted" by averaging with pain scores for nonpainful regions, then this will undoubtedly reduce the responsiveness of the pain item. The SSM-Symptoms subscale averages the responses to items about back pain, leg pain, and sensory, motor, and balance disturbances. If some of the items are not relevant to the patient, this averaging might render the subscale less responsive, in line with the "bandwidth *versus* fidelity" phenomenon.[32] Notably, the "neuroischemic" subdivision of SSM-symptoms was the least responsive of all the SSM subscales (Table 3). The consequence of attempting to assess every possible manifestation of a condition, rather than just its "quintessential" aspects, may paradoxically be "grasp all, lose all."

When considering the routine outcome assessment of all patients with lumbar spine problems within a hospital/practice, the ability to use a "one-size-fits-all" questionnaire considerably eases the administrative burden compared with using a condition-specific instrument for each separate disorder (*e.g.*, SSM for spinal stenosis, SRS-23 for spinal deformity, *etc.*). The availability and ease of administration of simple, brief instruments should encourage clinicians to collaborate with surgical registries.[33] Using one and the same instrument also allows for comparison between disorders, to gain an appreciation of the extent to which different spinal conditions impact on important aspects of the patient's life. On the down side, the use of a questionnaire such as COMI does not allow one to identify the specific aspect of function that remains impaired or improves after treatment, only that it still (or no longer) represents a problem to the patient.[34] The 12-month return rate was higher for the COMI than for the booklet containing the SSM and other questionnaires, although comparable efforts were made to obtain each of them. However, whether this was a function of the brevity of the COMI questionnaire or the added obligations of being in the observational cohort study[19] is not known.

This study was not without its limitations. The comparison of the COMI and SSM was not done using data collected within the same questionnaire booklet, as is usual for such studies. This might have led to somewhat lower correlations between the instruments.

An ongoing issue in all studies of questionnaire responsiveness concerns the external criterion used to indicate treatment success.[16,20,35–37] In this study we conducted the analyses using 2 different criteria: one (the GTO) that was included along with the COMI questionnaire and the other (SSM-sat), included with the SSM and other comparator questionnaires. Better correlations can be expected when the external criterion is included in the same booklet as the questionnaire under investigation. However, COMI showed similar (or even slightly better) responsiveness with SSM-sat as the external criterion as with GTO as the external criterion (and SSM-sat and GTO were also highly correlated) suggesting no such bias had occurred.

The proportion of patients in the "not successful" group was just 10% according to SSM-sat and 13% according to GTO. The small numbers may threaten the validity of the ROC analyses, which relied on these dichotomized outcomes. However, the correlations between the external criteria (using their 4- or 5-point response options) and the COMI/SSM change scores revealed similar findings to the ROC analyses for the rank order of the instruments' responsiveness (compare Tables 4 and 5), lending credibility to the findings. Nonetheless, the results should be verified in much larger groups of patients.

CONCLUSION

Although brief and not strictly condition-specific, the COMI did not seem to miss out on detecting important change in patients undergoing surgery for lumbar spinal stenosis. Indeed, it was as responsive as the condition-specific SSM. This provides further support for the use of the COMI in registries and clinical studies, with the added benefit of facilitating outcome comparisons with other spinal pathologies and reducing the response burden for the patient.

➢ Key Points

- ❑ The COMI is a brief, multidimensional patient-oriented questionnaire that is currently used in the routine assessment of outcome in patients with various spinal disorders.
- ❑ In patients with spinal stenosis, COMI change scores (preoperatively to 12 mo postoperatively) correlated to the hypothesized extent ($r \geq 0.4 \leq 0.8$) with the corresponding scores on the condition-specific SSM and its 2 main subscales (Symptoms and Physical Function), suggesting adequate longitudinal validity.
- ❑ Patient-oriented measures of treatment success 12 months after surgery correlated to a similar extent ($r = 0.5$–0.6) with the change scores (baseline to 12 mo) for the COMI and for the SSM and its main subscales.
- ❑ With treatment success as the external criterion, the area under the ROC curve was similarly high (>0.80) for the change score of the COMI and of the SSM and its main subscales.
- ❑ The COMI was well able to detect important change in patients with spinal stenosis and was as responsive as a condition-specific outcome instrument. The COMI is an effective instrument for use in spine surgical registries and clinical studies.

References

1. Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. *Spine* 1998;23:2003–13.
2. Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 2000;25:3100–3.
3. Mannion AF, Elfering A, Staerkle R, et al. Outcome assessment in low back pain: how low can you go? *Eur Spine J* 2005;14:1014–26.
4. Ferrer M, Pellise F, Escudero O, et al. Validation of a minimum outcome core set in the evaluation of patients with back pain. *Spine* 2006;31:1372–9; discussion 80.
5. Mannion AF, Porchet F, Kleinstück F, et al. The quality of spine surgery from the patient's perspective: Part 1. The Core Outcome Measures Index (COMI) in clinical practice. *Eur Spine J* 2009;18:367–73.
6. Mannion AF, Boneschi M, Teli M, et al. Reliability and validity of the cross-culturally adapted Italian version of the Core Outcome Measures Index. *Eur Spine J* 2012;21:S737–49.
7. Genevay S, Cedraschi C, Marty M, et al. Reliability and validity of the cross-culturally adapted French version of the Core Outcome Measures Index (COMI) in patients with low back pain. *Eur Spine J* 2011;21:130–7.
8. Damasceno LH, Rocha PA, Barbosa ES, et al. Cross-cultural adaptation and assessment of the reliability and validity of the Core Outcome Measures Index (COMI) for the Brazilian-Portuguese language. *Eur Spine J* 2011;21:1273–82.
9. Miekisiak G, Kollataj M, Dobrogowski J, et al. Cross-cultural adaptation and validation of the Polish version of the Core Outcome Measures Index for low back pain. *Eur Spine J* 2012;22:995–1001.
10. Storheim K, Brox JI, Lochting I, et al. Cross-cultural adaptation and validation of the Norwegian version of the Core Outcome Measures Index for low back pain. *Eur Spine J* 2012;21:2539–49.
11. Qiao J, Zhu F, Zhu Z, et al. Validation of the Simplified Chinese version of the Core Outcome Measures Index (COMI). *Eur Spine J* 2013;22:2821–6.
12. Roder C, Chavanne A, Mannion AF, et al. SSE Spine Tango–content, workflow, set-up. www.eurospine.org-Spine Tango. A European spine registry. *Eur Spine J* 2005;14:920–4.
13. Melloh M, Staub L, Aghayev E, et al. The international spine registry SPINE TANGO: status quo and first results. *Eur Spine J* 2008;17:1201–9.
14. Zweig T, Mannion AF, Grob D, et al. How to Tango: a manual for implementing Spine Tango. *Eur Spine J* 2009;18 (suppl 3):312–20.
15. Mannion AF, Kleinstueck FS, Mutter U, et al. The Core Outcome Measures Index (COMI) for the routine documentation of outcome in patients undergoing spine surgery: still performing well five years on?. Eurospine, Spine Society of Europe. Milan, Italy: *Eur Spine J* 2011:S441.
16. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Controlled Clin Trials* 1991;12(suppl):142–58.
17. Kirschner Guyatt A. A methodological framework for assessing health indices. *J Chronic Dis* 1985;38:27–36.
18. Stucki G, Liang MH, Fossel AH, et al. Relative Responsiveness of Condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *J Clin Epidemiol* 1995;48:1369–78.
19. Steurer J, Nydegger A, Held U, et al. LumbSten: the lumbar spinal stenosis outcome study. *BMC Musculoskelet Disord* 2010;11:254.
20. Mannion AF, Porchet F, Kleinstuck FS, et al. The quality of spine surgery from the patient's perspective: Part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. *Eur Spine J* 2009;18:374–9.
21. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
22. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 3rd ed. Oxford, United Kingdom: Oxford University Press Inc.; 1995.
23. Andresen EM. Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil* 2000;81:S15–20.
24. Husted JA, Cook RJ, Farewell VT, et al. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;53:459–68.
25. Hyland ME. A brief guide to the selection of quality of life instrument. *Health Qual Life Outcomes* 2003;1:24.
26. Bowling A. Just one question: if one question works, why ask several? *J Epidemiol Community Health* 2005;59:342–5.

27. Edwards P, Roberts I, Clarke M, et al. Increasing response rates to postal questionnaires: systematic review. *BMJ* 2002;324:1183.
28. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;4:293–307.
29. Fitzpatrick R, Davey C, Buxton MJ, et al. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 1998;2:i–iv, 1–74.
30. Abbott JH, Schmitt JS. The Patient-Specific Functional Scale was valid for group-level change comparisons and between-group discrimination. *J Clin Epidemiol* 2014;67:681–8.
31. Walsh TL, Hanscom B, Lurie JD, et al. Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. *Spine* 2003;28:607–15.
32. Cronbach LJ. *Essentials of Psychological Testing*. 5th ed. New York, NY: Harper and Row, 1990.
33. Balague F, Mannion AF, Pellise F, et al. Clinical update: low back pain. *Lancet* 2007;369:726–8.
34. Weiner BK, Walsh TL, Hanscom B, et al. Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. *Spine* 2003;28:607–15. *Spine* 2003;28:2304–5.
35. Beurskens AJHM, de Vet HCW, Köke AJA. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 1996;65:71–6.
36. Campbell H, Rivero-Arias O, Johnston K, et al. Responsiveness of objective, disease-specific, and generic outcome measures in patients with chronic low back pain: an assessment for improving, stable, and deteriorating patients. *Spine* 2006;31:815–22.
37. Havakeshian S, Mannion AF. Negative beliefs and psychological disturbance in spine surgery patients: a cause or consequence of a poor treatment outcome? *Eur Spine J* 2013;22:2827–35.
38. Roland M, Morris R. A study of the natural history of back pain. Part 1: Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983;8:141–4.
39. Exner V, Keel P. Erfassung der Behinderung bei Patienten mit chronischen Rückenschmerzen. *Schmerz* 2000;14:392–400.
40. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;37:53–72.
41. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. *Pain* 1986;27:117–26.